

Structural organization of the gene for human CD36 glycoprotein

Item Type	Journal article
Authors	Armesilla, Angel;Vega, Miguel A.
Citation	Armesilla, A.L. and Vega, M.A. (1994) Structural organization of the gene for human CD36 glycoprotein. Journal of Biological Chemistry, 269(29), pp. 18985-18991.
DOI	10.1016/S0021-9258(17)32263-9
Publisher	American Society for Biochemistry and Molecular Biology
Download date	2026-04-20 13:41:22
License	https://creativecommons.org/licenses/by/4.0/
Link to Item	http://hdl.handle.net/2436/7744

Structural Organization of the Gene for Human CD36 Glycoprotein*

(Received for publication, February 7, 1994, and in revised form, April 20, 1994)

Angel L. Armesilla[‡] and Miguel A. Vega[§]

From the Hospital de la Princesa and the Consejo Superior de Investigaciones Científicas,
C/Diego de León 62, Madrid 28006, Spain

The cell-surface glycoprotein CD36 interacts with a large variety of ligands, including collagen types I and IV, thrombospondin, erythrocytes parasitized with *Plasmodium falciparum*, platelet-agglutinating protein p37, oxidized low density lipoprotein, and long-chain fatty acids. Its expression is restricted to platelets, monocytes, adipocytes, and some endothelial and epithelial cells and is regulated during cell activation, differentiation, and development. CD36 belongs to a novel gene family of structurally related glycoproteins that includes CLA-1 and the lysosomal membrane glycoprotein LIMPII. To advance our knowledge on the genomic organization and the regulation of the cellular expression of the genes of this family, we have investigated the structural organization of the human CD36 gene and of its 5'-proximal flanking region. The CD36 gene is encoded by 15 exons that extend more than 32 kilobases on the human genome. Interestingly, the CD36 mRNA 5'-untranslated region is encoded by three exons. The 3'-untranslated region is contained in two exons, whose expression pattern can originate two mRNA forms. The cytoplasmic and transmembrane regions predicted at both terminal ends of the polypeptide chain are encoded by single exons, while the extracellular domain is encoded by 11 exons. The transcription initiation site of the CD36 gene is located 289 nucleotides upstream from the translational start codon. Sequence analysis of the proximal 5'-flanking region of the gene reveals the existence of a TATA box appropriately located with respect to the transcription initiation site and several potential *cis*-regulatory elements that might contribute to the transcriptional regulation of the CD36 gene. Delineation of the structural organization of the CD36 gene may help in defining the boundaries of relevant structural and/or functional domains in CD36 and, by extension, in the other members of the family.

CD36 is a cell-surface glycoprotein composed of a single polypeptide chain whose size ranges, depending on the cell type, from 78 to 88 kDa (50-kDa deglycosylated) (1, 2). CD36 glycoprotein is predicted to possess two transmembrane domains spanning residues 7–34 and 440–466, two short cyto-

plasmic tails at both the NH₂- and COOH-terminal ends (extending residues 1–6 and 467–472, respectively), and a large highly glycosylated extracellular domain comprising residues 35–439 (2, 3).

A multiplicity of interactions and functions have been ascribed to CD36. Briefly, CD36 acts as a receptor for the extracellular matrix glycoproteins thrombospondin (4–9) and collagen types I (10, 11) and IV (9). These interactions can participate in the modulation of phenomena such as platelet aggregation, platelet-monocyte interactions, retention of immature cells in bone marrow, and ingestion of apoptotic neutrophils by macrophages (4–12). CD36 mediates some of the adhesive properties of the erythrocytes parasitized with *Plasmodium falciparum* to the postcapillary venular endothelium (2, 13, 14). By this mechanism, parasitized erythrocytes evade their elimination in the spleen, facilitating red blood cell invasion and therefore parasite survival. CD36 has been implicated in the transport of long-chain fatty acids (15) and in the binding to oxidized low density lipoprotein (16). This low density lipoprotein-modified form has been shown to play a central role in the development of atherosclerotic lesions (17, 18). CD36 also interacts with the platelet-agglutinating protein p37, causing platelet aggregation. Such interaction may be of significance in thrombotic thrombocytopenic purpura (19, 20). Finally, physical association of CD36 with the *scr* family tyrosine kinases *fyn*, *lyn*, and *yes* (21) might explain the involvement of CD36 in signal transduction (22–25).

CD36 expression is restricted to platelets (where it initially was designated as glycoprotein IV), monocytes, some endothelial cells, mammary epithelial cells, adipocytes, activated keratinocytes, erythrocytes, and some tumor cell lines (1, 2, 13, 26–31). Expression levels of CD36 vary in some myeloproliferative disorders (32) and during the differentiation of some cell types. In this context, CD36 expression is induced during the differentiation of promonocytes to monocytes and macrophages (33), promegakaryocytes to platelets (33), and preadipocytes to adipocytes (15) and practically disappears during the differentiation of erythroblasts to erythrocytes (33). Thus, expression of the CD36 gene is subjected to both developmental and tissue-specific regulation.

CD36 and the recently characterized structurally related glycoproteins CLA-1 and LIMPII are members of a novel gene family (3, 34). The organization of the genes encoding these glycoproteins is not presently known.

Elucidation of the structural organization of the CD36 gene may be helpful for the definition of structural and/or functional domains of CD36 and related glycoproteins and will provide a starting point to unravel the molecular mechanisms controlling the tissue specificity of CD36 and its expression during cell activation and differentiation.

We have recently assigned the human CD36 gene to band q11.2 of chromosome 7 (35). In this paper, we describe the structural organization of the human CD36 gene, the identification of its transcriptional start site, and the structure of the proximal region of its promoter.

* This work was supported in part by Grant C086/91 from the Comunidad Autónoma de Madrid and Grant PM91/0138 from the Ministerio de Educación y Ciencia of Spain. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) Z32752–Z32765 and Z32770.

‡ Recipient of a predoctoral fellowship from the Ministerio de Educación y Ciencia of Spain.

§ To whom correspondence should be addressed: Hospital de la Princesa, Planta 9ª, C/Diego de León 62, Madrid 28006, Spain. Tel.: 34-1-3092115; Fax: 34-1-3092496.

EXPERIMENTAL PROCEDURES

5'-RACE-PCR

Single-stranded cDNA was synthesized at 42 °C for 1 h and 52 °C for 30 min from 4 µg of total RNA from the melanoma cell line C32 (13) using the CD36-specific oligonucleotide P1 (TGTGAAGTTGTCAGC-CTC). P1 spans nucleotides 618–601 of the published CD36 cDNA sequence (2). After removal of free oligonucleotide by centrifugation over Centricon 100, the sample was lyophilized and resuspended in 20 µl of water. 10 µl of the synthesized cDNA were poly(A)-tailed with 10 units of terminal transferase in the presence of 0.2 mM dATP at 37 °C for 5 min and 65 °C for 5 min. 1/100 of the above reaction mixture was added to a PCR mixture containing oligonucleotide P1 (antisense) and an oligo(dT)-adapter (GACTCGAGTGCAGATCGAT₁₇, sense); incubated at 94 °C for 2 min, 42 °C for 5 min, and 72 °C for 30 min; and amplified in a thermal cycler as follows: 94 °C for 30 s, 50 °C for 1 min, and 72 °C for 1 min for 35 cycles with a 2-s extension per cycle. 1 µl of the above PCR reaction mixture was amplified by PCR using primer P2 (GCTGCTGTTCATCATCAC, antisense, spanning nucleotides 453–436 of CD36 cDNA) and an oligo-adapter (GACTCGAGTGCAGATCG, sense) according to the following thermal cycle: 94 °C for 30 s, 50 °C for 30 s, and 72 °C for 2 min for 30 cycles with a 2-s extension per cycle. Amplified products were separated by gel electrophoresis, cut from the gel, and ligated to the pGEM-3T vector (Promega), and the mixture was used to transform competent *Escherichia coli* DH5α cells. Several colonies were randomly selected and sequenced.

3'-RACE-PCR

Single-stranded cDNA was synthesized at 42 °C for 1 h and 52 °C for 30 min from 2 µg of total RNA from the melanoma cell line C32 using the oligo(dT)-adapter (see above). 1/100 of the above reaction mixture was incubated at 94 °C for 2 min, 55 °C for 5 min, and 72 °C for 30 min in a PCR mixture containing oligonucleotide P3 (GACAACTATTGTTTCTGCACAG, derived from nucleotide fragment 1132–1153 of CD36 cDNA) and the oligo-adapter (see "5'-RACE-PCR"). This mixture was amplified as follows: 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 2 min for 35 cycles with a 2-s extension per cycle. 1/100 of the above PCR reaction mixture was again amplified by PCR using oligonucleotide P4 (TTCTGTATGCAAGTCCTG, derived from CD36 cDNA and spanning nucleotides 1247–1264) and the oligo-adapter according to the following thermal cycle: 94 °C for 30 s, 48 °C for 30 s and 72 °C for 2 min for 30 cycles with a 2-s extension per cycle. Amplified products were treated as described for 5'-RACE-PCR.

Isolation of Human CD36 Genomic Clones

Isolation of YACs Containing CD36 Gene—Screening of a human genomic DNA YAC library (36), generously provided by the Centre d'Etude du Polymorphisme Humain, was carried out by PCR using the conditions and reagents previously described for the identification of the chromosome that encoded the CD36 gene (35). Procedures to handle and isolate DNA from YACs were as described (37).

Screening of Human Genomic DNA Phage Libraries—The 5'-RACE-PCR fragment obtained as previously described (containing the 5'-untranslated region of the CD36 cDNA) and a PCR fragment amplified with a pair of oligonucleotides covering nucleotides 212–334 of the published CD36 cDNA sequence (2) were independently used as probes to screen 2 × 10⁶ phage clones of an *EcoRI*-digested human genomic library in λ phage Charon 21A and specific to chromosome 7 (ATCC 57722). CD36 cDNA (2) was used as a probe to screen 5 × 10⁶ phage clones of a human placenta genomic λEMBL3 library (CLONTECH). All hybridizations were carried out according to standard procedures.

S1 Nuclease Analysis—The region comprising nucleotides –267 to +43, relative to the transcriptional start site as determined by the 5'-RACE-PCR experiments, was obtained by PCR using oligonucleotides P5 (GAGCTCGAGTCTGACTTACTTGGATGGG, sense) and P6 (GGTCTCGAGGATCAAATGTTATTCGAGG, antisense). Oligonucleotide P5 spans nucleotides –267 to –249 of the 5'-flanking region of the CD36 gene. Oligonucleotide P6 was complementary to positions +43 to +23. Both oligonucleotides contained, at their 5'-ends, an *XhoI* recognition sequence preceded by 3 nucleotides to facilitate restriction enzyme cleavage. PCR was carried out with 20 µg of the λCh21ACD36.1 insert according to the following thermal cycle: 94 °C for 30 s, 52 °C for 30 s, and 72 °C for 30 s for 30 cycles with a 2-s extension per cycle. The resulting PCR products were cloned into the *XhoI* cloning site of the

pGL2-Basic vector (Promega). Clones were tested for the sense orientation and sequenced to verify the sequence of their PCR-derived inserts. This plasmid construct was designated as pGL2-CD36.267. A 340-nucleotide probe was synthesized by hybridizing ³²P-end-labeled oligonucleotide primer P7 (GATCAAATGGTATTCTGCAGG, corresponding to positions +23 to +43 relative to the transcriptional start site determined by 5'-RACE-PCR) to the denatured plasmid pGL2-CD36.267 and extending with Klenow DNA polymerase. Extended products were cleaved with the restriction enzyme *KpnI*, and the radiolabeled strand was separated from the template by alkaline agarose gel electrophoresis. 2 ng of the purified probe (with an activity of ~10⁶ cpm) were resuspended in 30 µl of 40 mM PIPES, pH 6.4, 1 mM EDTA, 0.4 mM NaCl, 80% formamide and hybridized overnight at 30 °C with 20 µg of total cellular RNA (38). After digestion with 200 units of S1 nuclease (Promega) for 30 or 50 min at 37 °C, protected fragments were analyzed on denaturing 8% polyacrylamide gels in parallel with a sequencing reaction of the λCh21ACD36.1 insert with primer P7.

Prediction of Transcription Factor-binding Sites—Prediction of putative transcription factor-binding sites was carried out by the C-coded program SITIOS (M. A. Vega), which includes Release 5.0 of D. Ghosh's transcription factors data base (39).

RESULTS

Determination of 5'- and 3'-Ends of Human CD36 mRNA by RACE-PCR—To delineate the complete structure of the CD36 gene, the 5'- and 3'-ends of its corresponding mRNA were determined by 5'- and 3'-RACE-PCR experiments, respectively, using the melanoma cell line C32 as the CD36 mRNA source (see "Experimental Procedures"). DNA sequencing of nine clones derived from the longest and most abundant 700-base pair band obtained as the result of the 5'-RACE-PCR experiments revealed that seven of them started at a common nucleotide, located 79 nucleotides upstream from the first nucleotide of the CD36 cDNA published sequence (2) and 289 nucleotides upstream from the translational start site (see Fig. 2). The other two clones started at the next 2 3'-adjacent nucleotides. Based on this result, nucleotide G, located 79 nucleotides upstream from the first nucleotide of the CD36 cDNA published sequence (2) and 289 nucleotides upstream from the ATG translational start codon, will be considered as the first nucleotide of the first exon of the human CD36 gene. This assignment was further confirmed by S1 nuclease analysis (see below).

3'-RACE experiments extended the 3'-end of the published CD36 mRNA by 24 nucleotides (2). This newly characterized region contained a polyadenylation signal located 21 nucleotides preceding the sequence C(A), which appears to be a preferred sequence for polyadenylation sites (see Fig. 2) (40).

Isolation of Genomic Clones Containing Human CD36 Gene—Initial screening of a human genomic DNA library in phage λEMBL3 with the whole CD36 cDNA as probe (2) led to the isolation of two partially overlapping positive clones. They were designated as λEMBL3CD36.8 and λEMBL3CD36.10 and contained inserts of 14 and 13 kb, respectively. Analysis of their inserts revealed that they contained the exons encoding nucleotides 410–1831 of the complete mRNA (Figs. 1a and 2).

To obtain the rest of the CD36 gene (that is, the exons encoding nucleotides 1–409 and nucleotide 1832 to the end), we first took into account that the CD36 gene is located on chromosome 7 (35). Therefore, an *EcoRI* λ Charon 21A library specific to chromosome 7 was screened with a 5'-RACE-PCR-derived probe. Two independent phage clones were isolated. Restriction fragment analysis of both of them revealed that they were identical. One of them, designated as λCh21ACD36.1 and which contained an insert of 5.7 kb, was chosen for further analysis. Restriction fragment analysis and DNA sequencing revealed that this clone included the first exon of the CD36 gene, comprising nucleotides 1 (the one determined by 5'-RACE-PCR) to 106 (Figs. 1a and 2). The screening of the same library with a PCR-derived fragment containing nucleotides 212–334 of the published CD36 cDNA (2) allowed us to isolate

¹ The abbreviations used are: RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; PIPES, 1,4-piperazinediethanesulfonic acid; kb, kilobase(s); YACs, yeast artificial chromosomes.

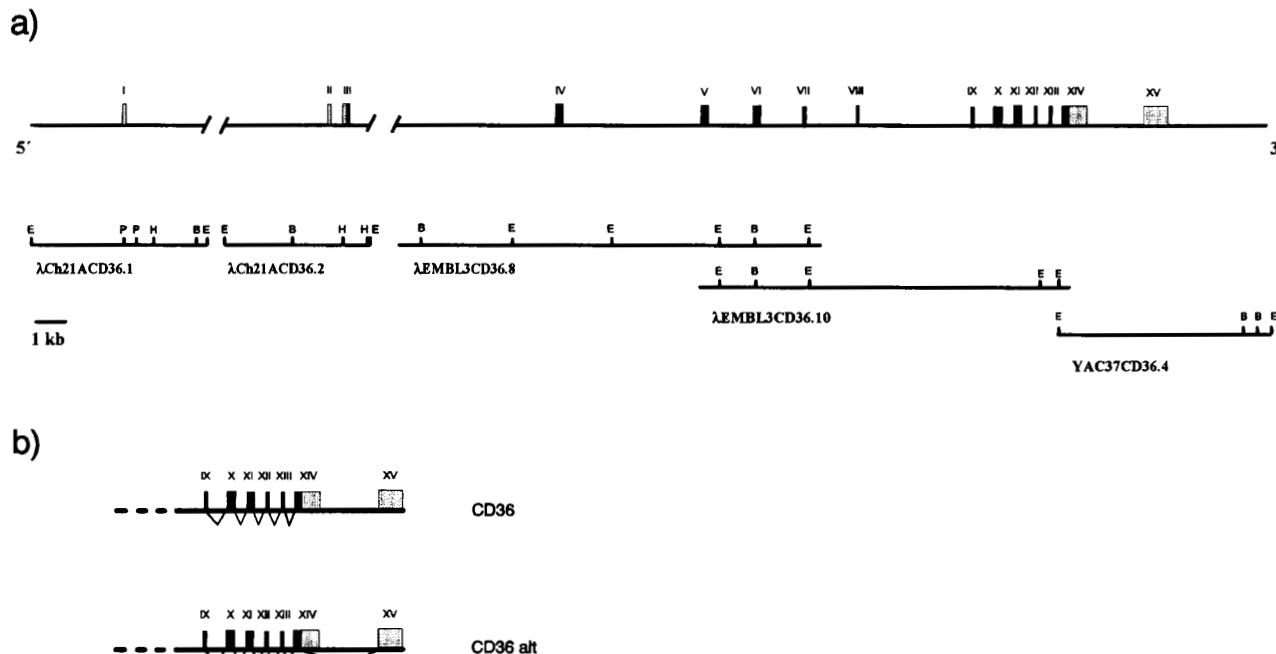


Fig. 1. *a*, physical map of CD36 gene. Exons are depicted as boxes numbered I–XV. Black boxes represent coding exons, and shaded boxes represent the 5'- and 3'-untranslated regions. Interrupted bars indicate noncontiguous DNA fragments. Regions spanning each of the genomic clones are indicated, as are their respective restriction maps. *E*, *EcoRI*; *B*, *BamHI*; *P*, *PstI*; *H*, *HindIII*. *PstI* and *HindIII* sites are only displayed for genomic clones λCh21ACD36.1 and λCh21ACD36.2. *b*, alternative mRNA forms of CD36. The drawing depicts the splicings that take place within the CD36 gene to originate the two mRNA forms of CD36, designated as *CD36* and *CD36 alt*, respectively.

one clone, designated as λCh21ACD36.2, that had an insert of 4.7 kb and that contained nucleotides 107–409 split into two exons (Figs. 1*a* and 2).

Since the 3'-untranslated region of CD36 was interrupted at the 3'-end of clone λEMBL3CD36.10, to search for the rest of the gene, we made use of one of the YACs containing the CD36 gene (YAC37) that was isolated as described under "Experimental Procedures." For this purpose, we first checked by Southern blotting the *EcoRI* fragment of YAC37 that hybridized to the 0.7-kb *EcoRI-SalI* fragment located at the 3'-end of the insert of clone λEMBL3CD36.10. We found out that the 3'-end of the gene was contained within a 7-kb *EcoRI* fragment. Therefore, a partial library was constructed in the plasmid vector pUCBM21 with YAC37 *EcoRI*-derived fragments ranging in size from 6.5 to 7.5 kb. The library was screened with the 0.7-kb *EcoRI-SalI* probe. Out of the 16 positive clones obtained, one clone, designated as YAC37CD36.4, was selected for further analysis. DNA sequencing revealed that it contained the probe used for the screening at one of its ends and uninterruptedly extended into the 3'-untranslated region of the CD36 cDNA down to the end of the mRNA (Fig. 1*a*).

Noguchi *et al.* (41) have recently isolated some cDNA clones from a variant of the erythroleukemia cell line K562 that displayed a different 3'-untranslated region of 623 base pairs that started immediately after the TAA stop codon. Analysis of clone YAC37CD36.4 revealed that this region was contained within an additional exon, located 1.9 kb downstream from exon XIV, and that it presented a consensus acceptor splicing site at its 5'-flanking region (Fig. 2). The nucleotide sequence of this new exon agreed with that reported by Noguchi *et al.* (41), except that we found 1 additional T nucleotide at position 425. The expression of this exon requires that the nucleotides in exon XIV immediately after the TAA stop codon act as an internal donor splicing site (in fact, its nucleotide sequence (GTAAGT) exactly matches the consensus sequence described for acceptor sites), removing the rest of the exon and joining the TAA sequence to the nucleotides of exon XV (Figs. 1*b* and 2). This

alternative mRNA form has also been found in two other leukemia cell lines, HEL and THP-1 (41). Whether its expression is cell type-dependent or whether it serves as a regulatory mechanism for CD36 expression is so far unknown. A physical map displaying the relative locations of the CD36 exons and the regions covered by each of the isolated and analyzed genomic clones is shown in Fig. 1*a*.

Structure and Organization of Human CD36 Gene—The organization of the CD36 gene was deduced from the analysis of different clones, each containing part of the gene. Such analysis included DNA sequencing basically on the exons, exon-intron boundaries, and 5'- and 3'-flanking sequences; extensive restriction mapping; Southern hybridization with different fragments and oligonucleotides derived from CD36 cDNA; and use of PCR to determine distances between adjacent exons. The results obtained from these experiments are summarized in Figs. 1*a* and 2 and Table I.

EcoRI fragments of the CD36 gene that included coding regions were equivalent to those identified by Southern blots of *EcoRI*-digested human DNA probed with CD36-derived probes (2). This observation is consistent with the existence of a single copy of the CD36 gene within the human genome.

The nucleotide sequence of all exons coincided with the published sequence of the cDNA (2), except that we found an insert of 4 nucleotides (GAAT) adjacent to the 3'-end of the termination codon, which is also found in the cDNA clones isolated by Noguchi *et al.* (41) (Fig. 2). The high degree of sequence conservation, as observed when the published CD36 cDNA sequence (2) was compared with the sequence deduced from the several genomic clones isolated, suggested that polymorphism of the CD36 gene was not extensive.

Analysis of the isolated CD36 genomic clones allowed us to conclude that the human CD36 gene consists of 15 exons and extends at least 32 kb on the q11.2 band of chromosome 7 (Figs. 1*a* and 2 and Table I) (35). Intron sizes vary from >5.9 kb (introns III and XIV) to only 0.5 kb (intron II). Most exons (V–XV) are contained within a region of <16 kb. Within it, two

cysteine cluster located at the half COOH-terminal region of the extracellular domain (2) was split into three exons with the following distribution: Cys-243 in exon VIII; Cys-272 in exon IX; and Cys-311, Cys-313, Cys-322, and Cys-333 in exon X. As occurred for the NH₂-terminal cytoplasmic and transmembrane domains, the COOH-terminal transmembrane and cytoplasmic regions were contained in a single exon (exon XIV), which also included a small portion of the extracellular domain.

All the splice acceptor and donor sequences conformed to the consensus rule for splicing (42). Of the 11 splice junctions found around coding exons, 46% occurred between codons (type 0), 18% occurred after the first nucleotide (type 1), and 36% occurred after the second nucleotide of a codon (type 2) (Table I). These data can be compared to the values of 41% type 0, 36% type 1, and 23% type 2 previously reported for vertebrate genes (43). Such a diversity in the type of intron phases limits the number of mRNA forms that might arise by exon deletion while maintaining the same open reading frame.

Determination of Transcription Initiation Site of CD36 Gene—To confirm the result of the 5'-RACE-PCR experiment with respect to the transcription initiation site of the CD36 gene, S1 nuclease analysis was carried out with a 340-nucleotide single-stranded ³²P-end-labeled antisense DNA fragment (see "Experimental Procedures"). After hybridization with total RNA from C32 or HeLa cells and S1 nuclease digestion, a small cluster of protected bands, centered around a size of 43 nucleotides, was observed only with RNA from CD36-expressing C32 cells (Fig. 3). The size of the most abundant protected band was in perfect agreement with the transcription initiation site identified by 5'-RACE-PCR. Taken together, 5'-RACE-PCR and S1 nuclease analysis demonstrate that transcription initiation of the CD36 gene starts at a single position located 289 nucleotides upstream from the ATG translational start codon. This result agreed with the presence of a consensus TATA box 28 nucleotides upstream from the transcription initiation site (see "Discussion").

Structure of Proximal Promoter Region of CD36 Gene—We have also determined the primary structure of the proximal region of the CD36 promoter. Computer analysis of the 5'-flanking region and the first exon of the CD36 gene revealed the existence of consensus sequences for promoter element TATA and CAAT boxes at positions -28 and -121, respectively, and for several known DNA *cis*-acting regulatory elements (Fig. 4; see "Discussion") (39).

DISCUSSION

This paper describes the structural organization of the gene encoding for the CD36 glycoprotein and the characterization of its transcription initiation site. Determination of the organization and structure of the CD36 gene provides the basis for investigation of the regulation, tissue specificity, expression, and pathogenic processes in which CD36 is involved.

The CD36 gene consists of 15 exons and extends at least 32 kb on the q11.2 band of human chromosome 7. A physical map displaying the relative locations of the CD36 exons and the regions covered by each of the isolated and analyzed genomic clones is shown in Fig. 1.

The delineation of the structure of the CD36 gene constitutes the first description of the genomic organization of a member of the gene family formed by the structurally related glycoproteins CD36, CLA-1, and LIMPII (34). Partial data support that CLA-1 and LIMPII genes may display genomic organization similar to that of the CD36 gene.² In this sense, we have reported the existence of a spliced form of CLA-1 (34). Assuming that CLA-1 possesses the same genomic structure as CD36, the

² D. Calvo and M. A. Vega, unpublished observations.

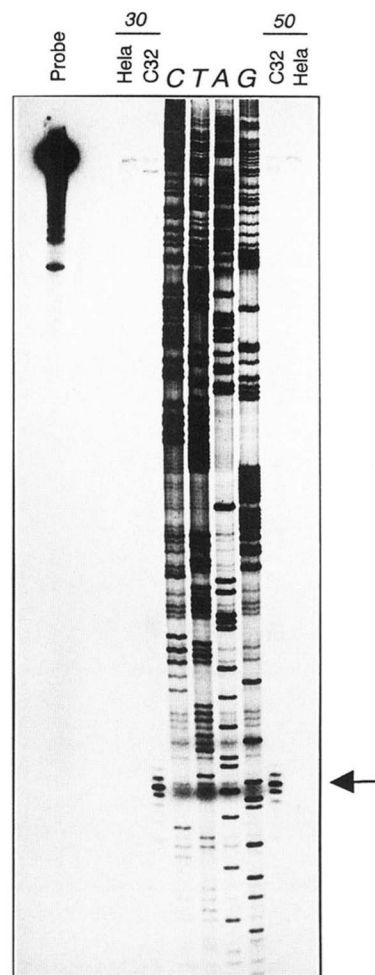


FIG. 3. Determination of transcriptional start site of CD36 gene by S1 nuclease analysis. S1 nuclease analysis was performed by annealing a 340-nucleotide single-stranded ³²P-end-labeled CD36 genomic probe to 20 µg of total RNA from C32 or HeLa cells and digesting the mixture with S1 nuclease for 30 and 50 min. Samples were analyzed on denaturing polyacrylamide gels. Undigested probe is shown on the left. The arrow points to the major protected fragment.

spliced form of CLA-1 would lack exons IV and V. Although mRNA forms arising by alternative splicing of coding exons have not been reported for CD36, a shorter CD36 mRNA form present in monocytes has been documented (44). In this regard, elucidation of the structural organization of the CD36 gene provides the structural basis to examine the existence of alternative splicing forms.

We have also identified a single major transcription initiation site for the CD36 gene. This site was appropriately located with respect to a consensus TATA box. Besides the regular translation initiation codon at position +290 (located in exon III), there is an ATG codon within the first exon at position +62 (Fig. 2). Translation starting at this position would originate a protein 19 amino acids long before reaching a stop codon located at position +119 in exon II. Such upstream codons are present in many protooncogenes (45). Removal of upstream ATG codons in the *lck* protooncogene activates translation and is associated with oncogenic transformation (46). Thus, the most upstream ATG sequence of the CD36 gene might play a role in translational control. Clearly, these observations deserve to be investigated.

A total of 11 exons (exons IV–XIII and part of exon XIV) encode the extracellular domain (Table I). The large number of exons encoding the extracellular domain may be a consequence of the multiplicity of interactions in which CD36 seems to be

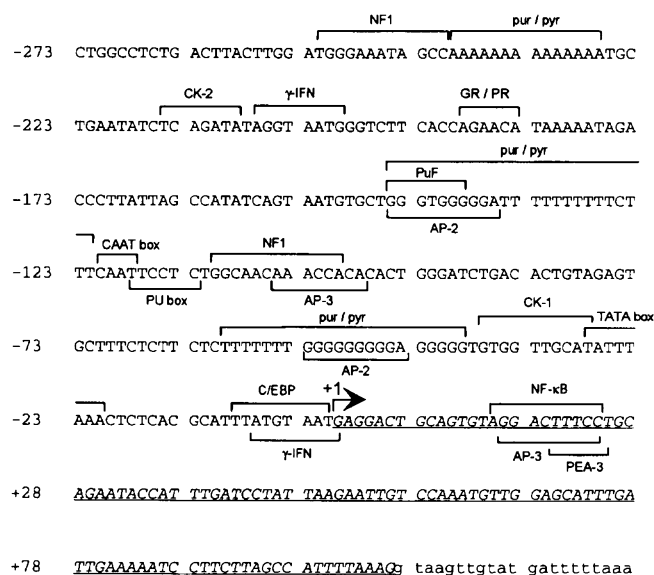


Fig. 4. Structure of 5'-proximal promoter region of CD36 gene. Nucleotides of the first exon are underlined and in upper-case italic letters. Nucleotides of the first intron are in lower-case letters. All numbering is relative to the transcriptional start site. The arrow indicates the direction of transcription. Marked regions correspond to consensus sequences for the indicated *cis*-acting regulatory elements. *pur/pyr*, purine/pyrimidine; γ -IFN, γ -interferon; GR/PR, glucocorticoid response/progesterone response; C/EBP, CCAAT/enhancer-binding protein.

involved (4–20) and suggests that the polypeptide chain may be organized into discrete portions that either alone or in combination can constitute independent structural and/or functional domains. In this regard, Leung *et al.* (8) have reported that amino acids 93–110 (encoded by exon V) and 139–155 (most of them encoded by exon VI) are implicated in the interaction of CD36 with thrombospondin. Interestingly, they have proposed that binding of region 139–155 to thrombospondin exposes CD36-binding site 93–110, resulting in a stable interaction with the ligand. On the other hand, the region spanning amino acids 97–110 (encoded by exon V) supports binding of CD36 to *P. falciparum*-infected erythrocytes (9). Altogether, the above results reinforce the proposal of the existence of discrete functional/structural domains in CD36 and highlight the importance of residues encoded by exons V and VI in some of the interactions mediated by CD36. Therefore, the elucidation of the exon-intron organization of the CD36 gene constitutes an interesting theoretical framework to investigate the structure-function relationship of this family of proteins by exon-shuffling experiments with the related CLA-1 and LIMPII genes.

Comparison of the CD36, LIMPII, and CLA-1 amino acid sequences revealed that human CD36 is more similar to rat CD36 than to human LIMPII and CLA-1.² Amino acid sequence identity between human and rat CD36 reaches 85% (15). When both sequences are compared exon by exon, identity ranges from 76% in exon XII to 93.5% in exon VII. Therefore, variability between human and rat CD36 appears dispersed all along the polypeptide chain. Practically the same variability pattern is obtained when amino acid sequences of human CD36, CLA-1, and LIMPII are compared (34).

We have also determined the nucleotide sequence of the proximal region of the CD36 promoter (Fig. 4). Although there is little information regarding the cellular factors that modulate CD36 expression, an abundance of transcription-binding sites was predicted to exist within the proximal region of the CD36 promoter, suggesting that a complex transcription scheme will control the expression of the CD36 gene. Besides

the TATA and CAAT boxes appropriately located with respect to the transcriptional start site (47), two ubiquitous NF1 sites were predicted at positions -111 and -252 (48). Phorbol ester-responsive elements AP-2 (at positions -53 and -145) (49), AP-3 (at positions +16 and -105) (50), a resembling NF- κ B/rel (at position +15) (51), and PEA-3 (at position +20) (52) may be involved in the regulation of CD36 mRNA expression by protein kinase C-mediated mechanisms. In fact, phorbol esters induce the expression of CD36 in myelomonocytic cell lines, while having the opposite effect on microvasculature endothelial cells (31). Induction of CD36 expression by γ -interferon in keratinocytes (29) and dermal microvasculature endothelial cells (31) may be dictated by the putative γ -interferon response elements (53) found at positions -7 and -207. The presence of several purine/pyrimidine-rich regions (sequences with a purine bias in one strand) found between nucleotides -60 and -39, -145 and -122 (this region included a purine factor site at position -145), and -240 and -227 (54, 55) may have special relevance in CD36 gene regulation. These regions of nuclease hypersensitivity occur predominantly in the 5'-region of genes and seem to have important roles in transcriptional regulation. Interestingly, a purine factor-binding factor is required for accurate transcription of the *c-myc* gene from its second promoter (54). The cytokine-specific sequences CK-1 and CK-2, located at positions -37 and -215, respectively, are found in the promoter sequences of some cytokine genes. Regulation of transcription of the granulocyte-macrophage colony-stimulating factor gene has been demonstrated upon binding of nuclear factors to these motifs (56). It is therefore conceivable that CD36 gene expression may also be controlled through similar mechanisms. The presence of a putative binding site at position -9 for transcription factors of the CCAAT/enhancer-binding protein family (57) would be consistent with the reported transcriptional effect that CCAAT/enhancer-binding protein factors have on the differentiation of preadipocytes to adipocytes (58) and promonocytes to monocytes (59) and therefore on the induction of CD36 expression during these processes (15, 33). The macrophage- and B cell-specific transcriptional activator PU.1 factor (60), predicted to bind at site -218, could contribute to the monocyte/macrophage expression of CD36. Glucocorticoids and steroid hormones, via the GR/PR response element predicted at position -189 (61), may exert important regulatory effects on the modulation of CD36 expression in macrophages and endothelial cells during inflammatory responses and in the control of the induction of CD36 expression in preadipocytes treated with dexamethasone (15). The role of the promoter region in determining transcriptional regulation and tissue specificity of the CD36 gene is currently under investigation.

Acknowledgments—We thank Dr. B. Seed for providing the CD36 cDNA and Dr. D. Le Paslier (Centre d'Etude du Polymorphisme Humain, Paris, France) for providing the human genomic YAC library. We also thank Drs. Angel L. Corbí, Dominica Calvo, and Arsenio Nueda for helpful discussions and Dr. Angel L. Corbí for critically reading the manuscript.

REFERENCES

- Greenwalt, D. E., Lipsky, R. H., Ockenhouse, C. F., Ikeda, H., Tandon, N. N., and Jamieson, G. A. (1992) *Blood* **80**, 1105–1115
- Oquendo, P., Hundt, E., Lawler, J., and Seed, B. (1989) *Cell* **58**, 95–101
- Vega, M. A., Segui Real, B., Garcia, J. A., Cales, C., Rodriguez, F., Vanderkerckhove, J., and Sandoval, I. V. (1991) *J. Biol. Chem.* **266**, 16818–16824
- Asch, A. S., Barnwell, J., Silverstein, R. L., and Nachman, R. L. (1987) *J. Clin. Invest.* **79**, 1054–1061
- McGregor, J. L., Catimel, B., Parmentier, S., Clezardin, P., Dechavanne, M., and Leung, L. L. (1989) *J. Biol. Chem.* **264**, 501–506
- Silverstein, R. L., Asch, A. S., and Nachman, R. L. (1989) *J. Clin. Invest.* **84**, 546–552
- Silverstein, R. L., Baird, M., Lo, S. K., and Yesner, L. M. (1992) *J. Biol. Chem.* **267**, 16607–16612
- Leung, L. L. K., Li, W. X., McGregor, J. L., Albrecht, G., and Howard, R. J. (1992) *J. Biol. Chem.* **267**, 18244–18250

9. Asch, A. S., Liu, I., Briccetti, F. M., Barnwell, J. W., Kwakye-Berko, F., Dokun, A., Goldberger, J., and Pernambuco, M. (1993) *Science* **262**, 1436-1440
10. Tandon, N. N., Ockenhouse, C. F., Greco, N. J., and Jamieson, G. A. (1991) *Blood* **78**, 2809-2813
11. Tandon, N. N., Kralisz, U., and Jamieson, G. A. (1989) *J. Biol. Chem.* **264**, 7576-7583
12. Savill, J., Hogg, N., Ren, Y., and Haslett, C. (1992) *J. Clin. Invest.* **90**, 1513-1522
13. Barnwell, J. W., Ockenhouse, C. F., and Knowles, D. M. (1985) *J. Immunol.* **135**, 3494-3497
14. Ockenhouse, C. F., Tandon, N. N., Magowan, C., Jamieson, G. A., and Chulay, J. D. (1989) *Science* **243**, 1469-1471
15. Abumratd, N. A., El-Maghrabi, M. R., Amri, E., López, E., and Grimaldi, P. A. (1993) *J. Biol. Chem.* **268**, 17665-17668
16. Endemann, G., Stanton, L. W., Madden, K. S., Bryant, C. M., White, R. T., and Protter, A. A. (1993) *J. Biol. Chem.* **268**, 11811-11816
17. Parthasarathy, S., Steinberg, D., and Witztum, J. L. (1992) *Annu. Rev. Med.* **43**, 219-225
18. Ross, R. (1993) *Nature* **362**, 801-809
19. Lian, E. C. Y., Siddiqui, F. A., Jamieson, G. A., and Tandon, N. N. (1991) *Thromb. Haemostasis* **65**, 102-106
20. Siddiqui, F. A., and Lian, E. C. Y. (1992) *Biochem. Int.* **27**, 485-496
21. Huang, M. M., Bolen, J. B., Barnwell, J. W., Shattil, S. J., and Brugge, J. S. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 7844-7848
22. Aiken, M. L., Ginsberg, M. H., Byers-Ward, V., and Plow, E. F. (1990) *Blood* **76**, 2501-2509
23. Ockenhouse, C. F., Magowan, C., and Chulay, J. D. (1989) *J. Clin. Invest.* **84**, 468-475
24. Trezzini, C., Jungi, T. W., Spycher, M. O., Maly, F. E., and Rao, P. (1990) *Immunology* **71**, 29-37
25. Schuepp, B. J., Pfister, H., Clemetson, K. J., Silverstein, R. L., and Jungi, T. W. (1991) *Biochem. Biophys. Res. Commun.* **175**, 263-270
26. Knowles, D. M., Tolidjian, B., Marboe, C., D'Agati, V., Grimes, M., and Chess, L. (1984) *J. Immunol.* **132**, 2170-2173
27. Kieffer, N., Beltaieb, A., Legrand, C., Coulombel, L., Vainchenker, W., Edelman, L., and Breton-Gorius, J. (1989) *Biochem. J.* **262**, 835-842
28. Greenwalt, D. E., Watt, K. W., So, O. Y., and Jiwani, N. (1990) *Biochemistry* **29**, 7054-7059
29. Juhlin, L. (1989) *Acta Dermato-Venereol.* **69**, 403-406
30. van Schravendijk, M. R., Handunnetti, S. M., Barnwell, J. W., and Howard, R. J. (1992) *Blood* **80**, 2105-2114
31. Swerlick, R. A., Lee, K. H., Wick, T. M., and Lawley, T. J. (1992) *J. Immunol.* **148**, 78-83
32. Clezardin, P., McGregor, J. L., Dechavanne, M., and Clemetson, K. J. (1985) *Br. J. Haematol.* **60**, 331-338
33. Edelman, P., Vinci, G., Villeval, J. L., Vainchenker, W., Henri, A., Miglierina, R., Rouger, P., Reviron, J., Breton-Gorius, J., Sureau, C., and Edelman, L. (1986) *Blood* **67**, 56-63
34. Calvo, D., and Vega, M. A. (1993) *J. Biol. Chem.* **268**, 18929-18935
35. Fernandez-Ruiz, E., Armesilla, A. L., Sanchez-Madrid, F., and Vega, M. A. (1993) *Genomics* **17**, 759-761
36. Albertsen, H. M., Abderrahim, H., Cann, H., Dausset, J., Le Paslier, D., and Cohen, D. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 4123-4127
37. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1987) *Current Protocols in Molecular Biology*, pp. 6.9.1-6.10.19, Green Publishing Associates, Inc./John Wiley & Sons, Inc., New York
38. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, pp. 7.66-7.70, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
39. Ghosh, D. (1991) *Trends Biochem. Sci.* **16**, 445-447
40. Fitzgerald, M., and Shenk, T. (1981) *Cell* **24**, 251-260
41. Noguchi, K., Naito, M., Tezuka, K., Ishii, S., Seimiya, H., Sugimoto, Y., Amann, E., and Tsuruo, T. (1993) *Biochem. Biophys. Res. Commun.* **192**, 88-95
42. Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459-472
43. Smith, M. W. (1988) *J. Mol. Evol.* **27**, 45-55
44. Lipsky, R. H., Sobieski, D. A., Tandon, N. N., Herman, J., Ikeda, H., and Jamieson, G. A. (1991) *Thromb. Haemostasis* **65**, 456-457
45. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125-8132
46. Marth, J. D., Overell, R. W., Meier, K. E., Krebs, E., and Perlmutter, R. M. (1988) *Nature* **332**, 171-173
47. Bucher, P., and Trifonov, E. N. (1986) *Nucleic Acids Res.* **14**, 10009-10026
48. Courtois, S. J., Lafontaine, D. A., Lemaigre, F. P., Durvieux, S. M., and Rouseau, G. G. (1990) *Nucleic Acids Res.* **18**, 57-64
49. Williams, T., and Tijan, R. (1991) *Genes & Dev.* **5**, 670-682
50. Chiu, R., Imagawa, M., Imbra, R. J., Bockoven, J. R., and Karin, M. (1987) *Nature* **329**, 648-651
51. Lenardo, M. J., and Baltimore, D. (1989) *Cell* **58**, 227-229
52. Faisst, S., and Meyer, S. (1992) *Nucleic Acids Res.* **20**, 3-26
53. Yang, Z., Suguwara, M., Ponath, P. D., Wessendorf, L., Banerji, J., Li, Y., and Strominger, J. L. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 9226-9230
54. Postel, E. H., Mango, S. E., and Flint, S. J. (1989) *Mol. Cell. Biol.* **9**, 5123-5133
55. Wells, R. D. (1988) *J. Biol. Chem.* **263**, 1095-1098
56. Shannon, M. F., Gamble, J. R., and Vadas, M. A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 674-678
57. Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T., and Kishimoto, T. (1990) *EMBO J.* **9**, 1897-1906
58. Vasseur-Cognet, M., and Lane, M. D. (1993) *Curr. Opin. Genet. Dev.* **3**, 238-245
59. Scott, L. M., Civin, C. I., Rorth, P., and Friedman, A. D. (1992) *Blood* **80**, 1725-1735
60. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., and Maki, R. A. (1990) *Cell* **61**, 113-124
61. Renkawitz, R., Schutz, G., von der Ahe, D., and Beato, M. (1984) *Cell* **37**, 503-510